*Comparative Effectiveness Review Disposition of Comments Report*

**Research Review Title:** *Diagnosis of Celiac Disease*

Draft review available for public comment from January 29, 2015 to February 27, 2015.

**Research Review Citation:** Maglione MA, Okunogbe A, Ewing B, Grant S, Newberry SJ, Motala A, Shanman R, Mejia N, Arifkhanova A, Shekelle P, Harmon G. Diagnosis of Celiac Disease. Comparative Effectiveness Review No. 162. (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2012-00006-I.) AHRQ Publication No. 15(16)-EHC032-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2016. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

## Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2175*
*Published Online: January 28, 2016*

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| Peer Reviewer 1 | Quality of Report | Good | No response necessary. |
| TEP 1 | Quality of Report | Good | No response necessary. |
| TEP 2 | Quality of Report | Superior | No response necessary. |
| Peer Reviewer 2 | Quality of Report | Superior | No response necessary. |
| Public Comment Dr. Barry Kisloff, American Gastroenterological Assn. | Quality of Report | The draft Diagnosis of Celiac Disease analysis that AHRQ presented for comment is a terrific source of relevant information regarding the current status of our ability (and limitations thereof) to diagnose Celiac Disease. In addition, it reflects that he believes the clinical evidence presented is accurate and that the information is presented clearly and completely and reflects the currently available clinical evidence. | No response necessary. |
| Peer Reviewer 3 | Quality of Report | Superior | No response necessary. |
| Peer Reviewer 4 | Quality of Report | Good | No response necessary. |
| TEP 3 | Quality of Report | Superior | No response necessary. |
| Peer Reviewer 1 | General Comments | The aim of this report is "…to assess the evidence on the comparative accuracy and possible harms of tests used for the diagnosis of CD, including serological tests, HLA typing, video capsule endoscopy, and endoscopic duodenal biopsy." This was further delineated into 4 key questions relating to 1) the comparative effectiveness of various serologies, genetic testing and capsule endoscopy on the accuracy of diagnosis, | Thank you. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | adherence, clinical outcomes and quality of life; 2) issues regarding the quality of biopsy performance and interpretation; 3) accuracy of diagnostic methods in select subsegments of the population; and 4) adverse effects of testing. The aims and methodology are clearly laid out and the results are largely in concordance with current guidelines and expert opinion. | |
| Peer Reviewer 1 | General Comments | It should be noted that certain relevant questions related to this review are not addressed in detail. Such questions include: 1) Can the duodenal biopsy be skipped? (This is briefly mentioned at the conclusion of the Executive Summary and at the end of the manuscript.) 2) Who should be tested for celiac disease, i.e. what symptoms and associated conditions should prompt testing? Should asymptomatic first-degree relatives be screened, and if so, when and how often? 3) What are the long-term implications of undiagnosed celiac disease? This is the inverse of the KQ4, the adverse effects of testing. It may benefit the reader to point out in a prominent spot (perhaps the Executive Summary) that these important questions remained beyond the scope of this report. | Thank you for the suggestions. These questions are highly relevant, but, as noted, were beyond the scope of the project. We hope that professional societies might use this systematic review as a springboard for policy recommendations or potential guidelines. |
| TEP 1 | General Comments | Yes, no comments about the scope and clinical meaningfulness of the review. | No response necessary. |
| TEP 2 | General Comments | The report does not address the crucial issue of what tests are needed to find the great majority of patients with celiac disease that remain undiagnosed. It will be relevant to decisions made by clinicians and by consumers in determining the most appropriate means for testing for celiac disease. It also provides evidence-based information helpful to practicing physicians, particularly gastroenterologists and pediatric gastroenterologists. | We agree. Unfortunately, the question of how to best find celiac disease in the general population was beyond the scope of this small systematic review. This report includes only studies where all subjects received both serology and duodenal biopsy. Only one study of screening in the general population met this inclusion criterion. We realize that cost and acceptance of biopsy by asymptomatic subjects might make such studies prohibitive; this is mentioned in the Discussion section. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| TEP 2 | General Comments | The key questions are appropriate and are explicitly stated. One particular question that perhaps is not answered directly is Key Question #3: Symptomatic Patients Vs Non-Symptomatic Individuals at Risk. The second component of the answer, for example, on Executive Summary page ES-12 (subsection of Key Question 3) in the table—the conclusion is insufficient. Evidence to address differences in serologic accuracy between patients with risk factors vs symptomatic population could not be determined. However, while there are very few comparative studies that look at accuracy of symptomatic and other at-risk groups, there are many stand-alone screening studies in patients with type 1 diabetes and iron deficiency anemia from which it is possible to extract information on diagnostic accuracy as many of these incorporated confirmatory biopsies for confirmation. There should be sufficient data to support a statement that serologic tests are highly accurate in patients with iron deficiency anemia or type 1 diabetes, despite there being no direct comparison within studies between these two groups. I wonder if it would be possible to compare the relative accuracy between papers rather than within studies for example that could help to provide some evidence to support at least the statement of accuracy in the asymptomatic individuals at risk. | We included all identified studies on accuracy of diagnostic tests in people with type 1 diabetes or iron deficiency that met our inclusion criteria, which required that all subjects underwent both serology and duodenal biopsy. This meant that only two studies of diabetics (Mansour, 2011 and Nevoral, 2013) and two studies of persons with iron deficiency anemia (Emami, 2012 and Cekin, 2012) were included. We do not feel this evidence is sufficient. However, please keep in mind that absence of evidence should not be interpreted as evidence of inaccuracy. We tried to make this point clear in the Discussion. |
| Peer Reviewer 2 | General Comments | This is a well written report which adds to the current information on diagnosis of celiac disease. This report provides useful and relevant information to both clinician and general population. Report correctly identified the key questions and they are stated clearly. | Thank you. No response necessary. |
| Peer Reviewer 3 | General Comments | This report summarizes a significant literature reports related to diagnosis of celiac disease. | Thank you. No response necessary. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | This has major implications for clinical practice, as it can help clinicians to identify the most appropriate strategy for diagnosis celiac disease in their patient population. | |
| Peer Reviewer 4 | General Comments | Very relevant study much needed. Would have liked more discussion about testing in the less than 3 yr old. Also, perhaps comparing different labs for the sensitivity and specificity. Some discussion about the costs of the different tests would have been great. | While important, cost issues were beyond the scope of this small systematic review. We have added mention of potential differences in accuracy among laboratories under "Applicability" in the Discussion section. |
| TEP 3 | General Comments | This is a well conducted review that essentially confirms what already known in literature, outlining strengths and limitations on studies focused on celiac disease diagnosis. | Thank you. No response necessary. |
| Peer Reviewer 1 | Executive Summary | [pg ES-10, Table A, Line 12] Results: Executive Summary Table A: It may be confusing to start KQ1 with video capsule endoscopy, as this is not a preferred or recommended testing modality except in selected circumstances. The reader may be under the impression that this test is on par with TTG testing. | Thank you for the suggestion. We agree with the reviewer; we moved video capsule endoscopy to after the serological tests throughout the report. |
| Peer Reviewer 1 | Executive Summary | [Table A, pg ES-12, line 41 and pg 49, line 43] KQ3: I am not certain that the conclusion that "both tTG and DGP tests tend to be less sensitive in adults than children" is entirely accurate. The discussion of children under 24 months versus older children does not include the recognized phenomenon of young children who test negative for TTG but positive for DGP who have a clinical and histological diagnosis of celiac disease; see for example Barbato, et al. Dig Liver Dis. 2011;43:465-9. This was not a direct comparison of children to adults, but it is widely recognized that TTG is relatively insensitive in children younger than 24 months. | Upon re-examination of the data, we agree. We have revised our conclusions accordingly. The report points out in several places (page ES-2, for example) that DGP tests may give a positive result in some individuals with CD who are anti-tTG negative, including children younger than two years. Barbato, 2011 is cited. However, only one study comparing accuracy between children under 24 months old and older children met our inclusion criteria. Olen, 2012, reported higher sensitivity and specificity (.96 and .98 respectively) for children under age 24 months compared to older children (.94, .86, respectively) for tTG IgA tests. They also reported higher accuracy for DGP IgA tests (sensitivity = 1.00, specificity = .31) in children under 24 months compared to in older children (sensitivity = .91, specificity = .26). |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| Peer Reviewer 1 | Executive Summary | [Table A, pg ES-13, line 20, and Table 23, pg 62, line 18] KQ4: The authors site a study showing higher sedation requirements during endoscopy in patients with celiac disease compared to controls.) I do not believe that this study is relevant to a discussion of harms related to endoscopy, as there were no mentions of sedation-related complications in this study; rather, that paper hypothesizes differences in physiology, comorbidity, or chronic home medication use (e.g. anxiolytics) among patients with celiac disease. | Thank you, we have removed the study from this review. |
| Peer Reviewer 1 | Introduction | The Introduction is well-written and appropriate. | Thank you. No response necessary. |
| Peer Reviewer 1 | Introduction | [pg ES-1, line 39 and pg 1, line 40] Re: the sentence (also in the Executive Summary) "A number of diagnostic methods have been developed, and their validity and acceptability remain controversial." This is not true of all diagnostic methods. TTG IgA is widely accepted by the medical community as an appropriate diagnostic test. Some tests, such as the DGP serologies and video capsule endoscopy, remain subject to debate. But it should be pointed out at the outset that there is little controversy about the utility of TTG IgA testing. | Thank you. We have revised to state that some of the more recent methods are controversial, including algorithms comprised of several tests. |
| TEP 1 | Introduction | [Pg 2, line 22] In the introduction, as well as a few other places, the review states that in cases of suspected IgA deficiency, IgG tTG should be ordered. In fact, the IgG tTG is rarely warranted as it is inferior to IgG DGP. I would suggest that this is stated somewhere and other references to IgG tTG are removed. In fact, the issue of celiac testing in known or suspected IgA deficiency should be included somewhere in this review as it is a frequent clinical concern in all age groups. | Your point is well taken. This review includes a section on diagnosis in IgA deficient individuals under Key Question 3.<br>We have removed the statement about ordering IgG tTG tests from the Introduction. |
| TEP 2 | Introduction | [Pg 2, line 14] The introduction is generally quite accurate. There is one point that is stated in the | Your point is well taken. We have removed specific data on accuracy from the Introduction. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | introduction but not supported by the data. The high sensitivity of endomysial antibody as being virtually 100% is contradicted by the data actually presented. Perhaps some rewriting or clarification of this section in the introduction to be more congruent with the actual data would be helpful. | |
| Peer Reviewer 2 | Introduction | This is clear, concise and introduces the which tests are relevant for diagnosis of celiac disease and how they have evolved over a period of time. | Thank you. No response necessary. |
| Peer Reviewer 3 | Introduction | I have no specific comments regarding the introduction. | No response necessary. |
| Peer Reviewer 4 | Introduction | Excellent | Thank you. No response necessary. |
| TEP 3 | Introduction | Properly crafted to put this review in context of the methodology used, allowing the reader to have full understanding of the approach taken to reach the conclusions outlined by the authors | Thank you. No response necessary. |
| Peer Reviewer 1 | Methods | The Methodology is sound and is described in appropriate detail. | Thank you. No response necessary. |
| TEP 1 | Methods | fine as written | Thank you. No response necessary. |
| TEP 2 | Methods | The special groups at risk do not mention family members. This is probably the group with the highest risk for celiac disease and the one for which screening is most often done. It is not clear that these particular studies are included in the risk assessment. Were family member studies of less than 300 excluded from analysis? | Along with persons with Type 1 diabetes, autoimmune disease, Turner's syndrome and Downs syndrome, family members were included as a population of interest under "Asymptomatic individuals at risk of celiac disease" in the list of PICOTs (populations, interventions, comparators, outcomes, timing). We included studies of family members with less than 300 subjects. However, only one study that met our inclusion criteria provided accuracy data specifically for those with family history (Nevoral, 2013). Results are presented in Chapter 3 on special populations.<br>Other identified studies of accuracy in subjects with family history of celiac disease did not meet our inclusion criteria as they did not biopsy seronegative subjects. In addition, we identified a very large (N> |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | | 20,000) proof of concept study using statistical learning methods and genome wide SNP profiles to predict celiac disease in subjects with family history; this method is experimental and outside the scope of our report, which focusses on methods used in current clinical practice. |
| TEP 2 | Methods | The overarching issue of how to best find celiac diagnosis in the general population is not addressed. What is the most effective way to find celiac disease in the overall population that includes symptomatic celiac disease, at-risk groups and asymptomatic celiac disease? It is the latter group that likely make up the great majority of celiacs that is left out. Indeed, an equally difficult question is how to identify the symptomatic group who should undergo testing. Do GI symptoms identify them? There are many studies in the United States, Mexico, and Europe that address the accuracy (see methods section). The report should at the very least address the accuracy of serology in the general population screening and perhaps point out the glaring gaps in our knowledge that could inform overall celiac disease detection strategies. | Unfortunately, the question of how to best find celiac disease in the general population was beyond the scope of this small systematic review. This report included only studies where all subjects, regardless of serology results, underwent duodenal biopsy. Only one study of screening in the general population met this inclusion criterion. We realize that cost and acceptance of biopsy by asymptomatic subjects might make such studies prohibitive; this is mentioned in the Discussion section. |
| TEP 2 | Methods | Was any account for a selection bias for serologic accuracy studies considered, in particular in geographic locations or populations where serologic tests are the most common initial detection test for celiac disease? Those patients are then referred for endoscopy where they undergo biopsies to confirm celiac disease, and then their serum samples are saved used either on a prospective or even a retrospective study for accuracy of a new serologic test. Could the prior selection bias because of a prior positive serologist test enrich the group with seropositive | The studies you describe are case-control studies where stored blood samples of biopsy-positive patients are used as "cases" of celiac disease in testing the accuracy of a new serologic test. These studies were relatively rare; but we did include some and their strengths and weakness are discussed. This type of selection bias was taken into consideration in our Strength of Evidence (SOE) ratings. The large number of high quality, non case-control studies of EmA, tTG, and DGP tests in symptomatic subjects led to high SOE ratings, while the small amount of studies, the poor quality |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | celiac disease? Might this explain the lower sensitivity in adults due to either having other causes of villous atrophy or perhaps more adult patients having their celiac disease discovered by scoping and biopsy first? | (including case-control design), inconsistent results, led to a rating of "insufficient" SOE for accuracy of these tests in asymptomatic subjects in specific risk groups. |
| TEP 2 | Methods | Otherwise, the inclusion/exclusion criteria are quite justifiable. The search strategies are clearly stated and logical. The definitions or diagnostic criteria for outcome measures are appropriate and the statistical methods seem robust. | Thank you. No response necessary. |
| Peer Reviewer 2 | Methods | This is a sound and complete report. I find appropriate inclusion and exclusion criteria were used. Search strategies were clearly stated and easy to understand and follow. Definitions and statistical methods were appropriate. | Thank you. No response necessary. |
| Peer Reviewer 3 | Methods | All the inclusion and exclusion criteria were clear. I have no concerns regarding the statistical methods. | Thank you. No response necessary. |
| Peer Reviewer 4 | Methods | yes ,absolutely | Thank you. No response necessary. |
| TEP 3 | Methods | Standard methodology that is appropriate for this kind of studies. | Thank you. No response necessary. |
| Peer Reviewer 1 | Results | [pg 18, line 5, column 3] Table 7 header: change "Perfect Female" "Percentage Female" | Good catch. We have revised accordingly. |
| TEP 1 | Results | As written, one could take away the concept that VCE is similarly useful and well validated as tTG or EMA. This is not a correct conclusion. The VCE data is limited and based on studies with high pretest probability of celiac disease and from a few centers with specialization in VCE. In most regions VCE would likely perform much less well, although there is limited data either way. | We agree that VCE is not as useful as serology and has far less evidence of accuracy. As the Executive Summary (ES) and the Conclusions indicate that the SOE (Strength of Evidence) is rated high for both tTG and EmA, while rated moderate for VCE, we hope that the reader would not take away a message that VCE is similarly useful or valid. We think it is clear from the Summary Table that there is far less evidence on VCE.<br>We tried to make clear in the Introduction and ES that VCE is not a traditional means of detecting celiac disease and is only used for adults who wish to avoid biopsy. We have re-organized the entire |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | | report so that VCE is always discussed after the first line methods of diagnosis. |
| TEP 1 | Results | The section on gluten challenge is also misleading. While 2 weeks of gluten challenge can induce intestinal changes in many adults, a substantial proportion will require longer duration of gluten challenge, similar to children. The recent ACG algorithm on this subject is more correct. Also 15 grams of gluten per day is not supported by recent literature which suggests 3 or more grams of gluten per day is generally sufficient. | Your points are well taken. We have revised to include the ACG algorithm (Rubio-Tapa, 2013) that recommends an additional six weeks of gluten challenge for adults who can tolerate. We also include your recent study which found that 3 grams is sufficient (Leffler, 2012). |
| TEP 1 | Results | Finally, I think the conclusion that serologies are more accurate in children compared with adults is an overstatement. There is a great deal of heterogeneity in age cutoffs, tests used and populations. In general results are comparable within the limitations of the studies. | Thank you. We have downgraded the strength of evidence for this finding to low. |
| TEP 2 | Results | The amount of detail presented is appropriate. Characteristics of studies are clearly described and the key messages provided are relevant quite clear. The tables, figures and appendices are adequate and quite descriptive. | Thank you. No response necessary. |
| TEP 2 | Results | General Population Screening<br>One area of study that is alluded to in the conclusions, but is not really addressed in terms of accuracy are the large number of studies that have been done in screening populations for celiac disease using an initial serodiagnostic approach . There are probably 100 such studies and some of these are quite large in number and the usual design is to screen a general population and then invite the positive patients for endoscopy for biopsy confirmation. Whilst this cannot address the sensitivity of the serologic tests, it could be useful to determine the specificity of positive serologic tests in the context of the | Unfortunately, the question of how to best find celiac disease in the general population was beyond the scope of this small systematic review. This report included only studies where all subjects received both serology and duodenal biopsy. Only one study of screening in the general population met this inclusion criterion. We realize that cost and acceptance of biopsy by asymptomatic subjects might make such studies prohibitive. We mention in the Research Gaps section. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | general population. There are a few studies that have parallel endoscopy and serology done in which neither the results of endoscopy nor the results of serology determined the likelihood of undergoing biopsy. Perhaps inclusion of these studies may be helpful or else should be explicitly excluded. This fairly extensive literature would help address the accuracy of serologic testing in patients without symptoms and perhaps even those not belonging to any at-risk group in the general population. | |
| TEP 2 | Results | There is data which is referenced in the special populations group, particularly IgA deficiency, regarding IgG DGP tests. However, this particular test has been incorporated into many larger symptomatic populations and shows a high level of accuracy. Perhaps this discussion should be included under the Key Question #1 in addition to the accuracy of IgA DGP to add in also the accuracy of IgG DGP as this test may have relevance in the IgA sufficient population, not simple relegated to use in the IgA deficient population. | Thank you. We have added results on accuracy of IgG DGP tests in IgA sufficient patients to Key Question 1. We cite a new systematic review (Collatz Schyum, 2013) which reports on seven studies of IgG DGP in IgA sufficient adults; |
| TEP 2 | Results | The testing in the IgA deficient population. There are a couple of studies—one a large Italian study in patients with IgA deficiency that might be relevant to studying the accuracy of serologic testing in IgA deficient individuals. | Thank you. We have added two new studies (Bienvenu, 2014; Wolfe, 2014) to this section. However, no large Italian studies met our inclusion criteria. We identified several studies by Picarelli and colleagues; these were excluded as seronegative subjects did not undergo biopsy. |
| TEP 2 | Results | Page 22 of 232. Safety of endoscopic biopsies. Whilst there may not be any literature on the safety of endoscopic duodenal biopsies in patients with celiac disease, there is likely some literature on the safety of upper endoscopy and duodenal biopsies in the general population that may be very relevant to patients undergoing endoscopy with celiac disease. Perhaps inclusion of a review | We found no safety data specific to persons undergoing these procedures for celiac disease investigation. While an exhaustive review on the safety of upper endoscopy and duodenal biopsy is beyond the scope of this small review, we have added a brief summary of adverse events in the general population. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | of this literature would be informative. | |
| TEP 2 | Results | Page 25 of 232. Marsh classification is not used by most clinical pathologists. Most would use a simple qualitative assessment of villous architecture and whether or not intraepithelial lymphocytes are elevated or not. Marsh has been used in some research studies. There are other competing classification systems, including the Villanacci score. | Thank you, we are aware that Marsh classification is used primarily by researchers rather than clinicians. We have added language to the Discussion section to make clear that although Marsh classification is used in the vast majority of studies on diagnostic accuracy, it is not often used in clinical practice. |
| TEP 2 | Results | Page 25 of 232. Endomysial antibody is not 100% sensitive and the introduction to the main body as well as the background of the executive summary should qualify the description of how accurate the endomysial antibody is to reflect their actual data. | Thank you. We have removed the data on accuracy of EmA test from the Introduction sections. |
| TEP 2 | Results | Page 55 of 232. The difficulty in comparing threshold for tTG antibodies is the fact that different test kits use different reference ranges by design. Some of these kits have adjustment of the kit manufacturer reference ranges by individual testing laboratories make these even more difficult to normalize across studies. This is perhaps one of the significant gaps in the research and something that hampers the comparability of test strategies. There are also some ongoing proficiency surveys—one by the College of American Pathologists and another in the U.K.—that do compare the performance of testing and the accuracy in standardized serum samples. In particular, the U.K. study has shown substantial variability in the quantification between laboratories. In addition, there are several studies that have examined variation between labs and in testing. | Thank you. We now mention this issue under Applicability in the Discussion section and cite studies comparing accuracy among different laboratories and manufacturers. |
| TEP 2 | Results | Page 69 of 232. The question of challenge and the duration of challenge. IgA deposits in biopsies are not used clinically in North America or Europe | Your points are well taken. We have revised this section to include the ACG algorithm (Rubio-Tapa, 2013) that recommends at least two weeks of gluten |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | outside of a few research labs. There is one study completed in North America of a small group of patients with treated celiac disease who were challenged with moderate doses of gluten for shorter periods by Leffler et al (published in Gastroenterology) that may be informative on dose/duration. | challenge before conducting biopsy, and an additional six weeks for adults who can tolerate. We also include in the results Leffler's study finding that 3 grams is a sufficient dosage (Leffler, 2012). |
| Peer Reviewer 2 | Results | Studies analyzed and outcomes were well described. For the serological assays it clearly brought out the fact that different laboratories may use different cut off, however the sensitivities and specificity are in the same range. One issue that can be additionally discussed is the different HLA DQ2 haplotypes and methods for HLA typing. Report clearly states that absence of HLA-DQ2/8 rules out the diagnosis of celiac disease. However, there are two major HLA DQ2 haplotypes: 1. HLA-DQ2.5: HLA-DQA1*05: XX-HLA-DQB1*02:01 and 2. HLA-DQ2.2: HLA-DQA1*02: XX-DQB1*02:02 and single copy of HLA-DQ2.5 is strongly associated with celiac disease while clinical significance of a single copy of HLA DQ2.2 is unknown. One major study about this is by Pietzak MM et al, Clin Gastroenterol Hepatol. 2009 Sep;7(9):966-71. Such HLA-DQ haplotypes can only be identified when using molecular method for HLA typing but not by serological method for typing. | Thank you so much for your comment. Unfortunately, HLA typing beyond serologic tests for HLA-DQ2 and HLA-DQ8 halotypes was beyond the scope of this small project which covers only diagnostic methods currently used in clinical practice in the US.. Dr. Pietzak practices nearby and was a member of our technical expert panel, so we are aware of the important study you mentioned. As noted, the other HLA-DQ halotypes can only be identified using molecular method; this is not current clinical practice. |
| Peer Reviewer 3 | Results | All the data is presented clearly. | Thank you. No response necessary. |
| Peer Reviewer 4 | Results | the figures very very busy -too much data in one table | Thank you. We realize that the figures are busy, but we feel this is the best way to present the accuracy results for each test. Several expert panel members and peer reviewers commented that the tables and figures were user friendly. |
| TEP 3 | Results | The tables presented are extremely useful to guide the reader through the review process that | Thank you. No response necessary. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | led to the results presented. Very helpful. | |
| Peer Reviewer 1 | Discussion/Conclusion | [pg v, line 43] The conclusion of the Abstract section states "Additional studies are needed to increase the strength of evidence of accuracy of diagnostic tests in special populations and to validate promising algorithms." This would benefit from more practical advice based on the report's findings. For example, the following paraphrase of the report's conclusion would be an appropriate conclusion: "New evidence on accuracy of tests used to diagnosis celiac disease supports the high sensitivity of IgA tTG tests and high specificity of both IgA tTG and IgA EmA tests reported in recent systematic reviews. High strength of evidence of accuracy was found for IgA DGP tests but accuracy is slightly less than that of the other serological tests." | Thank you. We have revised the abstract per your comment. The strength of evidence for DGP was re-classified as moderate. |
| Peer Reviewer 1 | Discussion/Conclusion | [pg v] In the Abstract, consider including the caveat that the sensitivity of serologies is dependent on the subject maintaining a gluten-containing diet at the time of testing. | While we agree this is important, word count limitations require us to focus on the most important points. The Executive Summary and body of the report make clear that one must maintain a gluten-containing diet at time of testing. |
| TEP 1 | Discussion/Conclusion | no major concerns | Thank you. No response necessary. |
| TEP 2 | Discussion/Conclusion | The implications are clearly stated, and limitations are also adequately described. The discussion includes mention of asymptomatic screening studies, but really does not review the extensive literature in this area (see above). The future research section does point out significant gaps in our knowledge base and gives clear direction as to research directions that should be incorporated. One issue that is not mentioned is the need for even more accurate serologic tests that could supplant the need for biopsies in most patients. In addition, the ESPGHAN criteria for biopsy avoidance applies only to patients with symptoms | As stated above, this report included only studies where all subjects received both serology and duodenal biopsy. Only one study of screening in the general population met this inclusion criterion. We realize that cost and acceptance of biopsy by asymptomatic subjects might make such studies prohibitive; we mention in the Research Gap section. |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | of celiac disease and when a second blood sample separately taken from the patient confirms the presence of endomysial antibody and carriage of the necessary HLA type. | |
| Peer Reviewer 2 | Discussion/Conclusion | Review is very inclusive and succinctly states the relevant findings. Discussion does bring out the differences in methods for performing the serological assays and the role of clinicians practice when performing biopsies. Additional discussion on HLA haplotypes and methods for HLA typing will be helpful. | As noted above, HLA-DQ halotypes other than DQ2 and DQ8 can only be identified using molecular method; this is not current clinical practice. |
| Peer Reviewer 3 | Discussion/Conclusion | The sensitivity of the EMA test is described in the report. In the "Key Points" on page 49 (lines 22-25), it is stated that "IgA EmA tests have lower sensitivity and similar specificity of IgA tTG tests". However, at another point in the report (example pg 25, lines 14-15) it is stated that "Almost 100% of patients with active celiac disease…will have the IgA class of anti-EMA antibodies". This is somewhat mis-leading. What exactly is meant by "active celiac disease"? Although it may be true that the majority of patients with active celiac disease (on gluten-containing diet, evidence of partial/total villous atrophy) may be EMA positive, this is not likely to carry into routine use. Testing for celiac disease is ordered on patients with a wide range of clinical symptoms; some of these patients may have only mild villous atrophy. In this group of patients, it is likely that the sensitivity of EMA will be lower than in the "active" celiac disease group. I just would not want to give the impression that EMA has such a high sensitivity, when in reality it appears to be lower than that of TTG-IgA. | We have removed the language regarding the percentage of active celiac disease patients with positive IgA EmA, along with all other serology data, from the Introduction (former page 25). The final report contains accuracy numbers only in the Results and Discussion sections of the Executive Summary and full report. |
| Peer Reviewer 4 | Discussion/Conclusion | very good<br>some more detailed discussion about serology positive biopsy negative patients would be great | Thank you. Video capsule endoscopy (VCE) is recommended as a replacement for biopsy when the patient cannot or will not have a biopsy. It is not |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | and would capsule help in those situations? | recommended as a secondary screening test when biopsy is negative and we identified no studies of this use. |
| TEP 3 | Discussion/Conclusion | It would be desirable that the authors would outline some of the major limitations and challenges in celiac disease diagnostic tools. Specifically:<br>1. If the intestinal biopsy is used as gold standard to establish sensitivity, specificity, PPV, and NPV, only studies starting with intestinal biopsy followed by serology analysis should be considered. Otherwise, it would be a "self-serving prophecy" to consider algorithms in which serology is followed by endoscopy (this approach would not allow false negative results for any given serological screening test considered. While the inherited invasive nature of an endoscopy is acknowledge, this poses problems, particularly for those extra intestinal manifestations of the disease (including dermatitis herpetiformis, anemia, joint pain, just to name a few) for which the procedure is not clinically indicated. | This systematic review included accuracy studies only if biopsy was performed regardless of serology results. All subjects in a study, both sero-negative and sero-positive, had to undergo biopsy. This allows calculation of a false negative rate. Studies that only biopsied subjects with positive serology for celiac disease were excluded from this report, for the very reason you stated. A list of these excluded studies is included in Appendix B.<br>In addition, in order for a study to meet our inclusion criteria, the authors and physicians performing the biopsy and interpreting the results had to be blinded from the results of serology. |
| TEP 3 | Discussion/Conclusion | 2. The finding that the intestinal biopsy can be not "gold" in community hospitals compared to academic hospitals should be further emphasized. Indeed, while lab test are objective measures, intestinal biopsies are at the mercy of subjective interpretations by pathologist. This report outlines also the inter-observed variation that cast doubts on the robustness of intestinal biopsy as gold standard | Thank you. This issue is now discussed in the Executive Summary and Discussion section under “Applicability.” |
| Peer Reviewer 1 | Clarity and Usability | The report is well structured and organized and the main points are clearly presented. | Thank you. No response necessary. |
| TEP 1 | Clarity and Usability | no major concerns | Thank you. No response necessary. |
| TEP 2 | Clarity and Usability | This is a well-structured and organized with main points clearly presented and will be used to inform | The Executive Summary and the Discussion section of the full report state the following regarding biopsy |

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| | | policy and practice decisions. A more explicit statement on biopsy avoidance would be helpful if indeed the data is convincing either way. | avoidance: "Notably, current European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) guidelines state that if a patient demonstrates a tTG result greater than (10x) the normal limit, then the patient will undergo an EmA test and HLA typing - if the patient tests positive, then responds to gluten exclusion diet, a diagnosis of celiac disease can be made without use of biopsy. These guidelines have not been adopted by societies in the U.S. at this time. Evidence seems to support that a multiple-testing strategy without biopsy is accurate; however, additional studies are needed to confirm which threshold levels and specific populations would benefit from increased accuracy." |
| Peer Reviewer 2 | Clarity and Usability | Report is well structured and organized. Main points are clearly summarized in the executive summary. This document will help a general physician whose primary practice does not specialize in celiac disease but most likely it the first to come across a patient. | Thank you. No response necessary. |
| Peer Reviewer 3 | Clarity and Usability | The report will provide information to clinicians that, I believe, will change the diagnostic approaches to patients with suspected celiac disease. | Thank you. No response necessary. |
| Peer Reviewer 4 | Clarity and Usability | once again the tables seem very busy | Thank you. We realize that the figures are busy, but we feel this is the best way to present the accuracy results for each test. Several expert panel members and peer reviewers commented that the tables and figures were user friendly. |
| TEP 3 | Clarity and Usability | The report is extremely clear and, therefore, accessible to average readership. | Thank you. No response necessary. |